Constantin-Renato IVĂNESCU

Constantin-Renato IVĂNESCU

Medical Data Processing Enhancements:
New AI-based Computation and Communication Models

Editura UNIVERSITARIA
Craiova, 2024

Referenți științifici:

Prof.univ.dr. Marian-Cristian Mihăescu
Conf.univ.dr. Smaranda Belciug

Descrierea CIP a Bibliotecii Naţionale a României
IVĂNESCU, CONSTANTIN RENATO
Medical data processing enhancements : new AI-based computation and communication models / Constantin Renato Ivănescu. - Craiova : Universitaria, 2024
Conţine bibliografie
ISBN 978-606-14-2105-3
61
004

1. Introduction

1.1. Problem Statement. Motivation and objectives

Medicine followed different stages throughout history, from the ancient Egyptians to modern medicine assisted by CAMD (Computer-Aided Medical Diagnosis). CAMD is a technology that uses elements of digital image processing and Artificial Intelligence (AI). Although AI has been around for decades, only recently has it paved the way for enhancing medical data processing. Nowadays, it has a great influence on medicine, heavily influencing healthcare professionals' activity.

AI was first introduced as an academic discipline in 1956 [1]. Many scientists have raised the possibility of creating an artificial brain since 1940. It should be mentioned that even if AI existed for more than 60 years, it is only now attracting all the attention and the reasons are various. One of them is the new parallel processing that GPUs (graphic processor units) offer at a low cost for training deep learning models. Another reason is the newly developed algorithms that are very efficient and can be merged with the GPUs to enable quicker and more accurate computations. The last one is the data that is generated by the medical system. There is a huge amount of data generated which needs to be processed in order to derive insights from it. In essence, this is artificial intelligence helping us to classify or make smart decisions [2].

In the healthcare system, artificial intelligence is used together with biochemical and/or clinical information from medical analyses. Briefly, CMD refers to analyzing digital medical images using software in order to determine how likely it is that a feature corresponds to a particular disease condition (e.g. malign vs benign).

In clinical practice, diagnosing a patient is a procedure that involves patient examination, medical tests, medical records, the patient’s medical history, etc. In the context of the rapid development of technology and the huge amount of available medical investigation, even the clinical examination can be subjective as clinical judgment is used by medical personnel to assess patients. Most of the time, a list of possible diagnoses is made by the doctors from which they will choose one depending on the patient's interview or history.

The revolution in AI in medicine fascinated and benefitted us. We have to take advantage of the evolution of technology and use as many possible non-invasive methods for patients. Intelligent CAMD systems use AI and have the capability to infer new knowledge by analyzing medical data. In this way, such systems can improve their performance together with the current diagnostic rules. They also must have a feedback mechanism.

There is a wide variety of diseases that require improvements in diagnosis and treatment and that can be addressed with CAMD [3][4]. In this work, however, we focus on different types of cancer like colon cancer, lung cancer, and breast cancer. For these diseases, there are some techniques used which are non-invasive and have high accuracy. A crucial task in the medical field is image analysis. It represents a basic modality for diagnosing any disease and moreover, image acquisition does not harm the human body. Image analysis is now dominated by AI.

In many hospitals, the staff can not correctly interpret the output of some medical techniques which represents a reason that has led to the development of computer-aided diagnosis. This reason derives from factors such as medical diagnosis process complexity, large amounts of diagnostic knowledge, and a large amount of clinical data. In this case, an important role can be played by a Machine Learning algorithm or a system of Machine Learning algorithms which represents one of the most important instruments in AI. These entities can be trained on the databases to assist the medical staff's decision-making process. Using such a system can help reduce costs that must be allocated for the devices and the costs for medical staff specialization. In this scenario, a bigger problem that can be solved is the subjectivity in interpreting medical images.

Nowadays, there is a need for high-speed decision-making systems in the medical domain where a huge amount of data is generated so new models must be created to analyze complex data [5]. ML techniques, along with the huge potential for cost reduction, can optimize patient management. It can make its own decision, we only have to feed data into it. We can associate the need for machine learning with:

- Solving complex human problems
- Decision making improvements
- Finding hidden patterns and trends in data
- Building statistical models
- Data generation from electronic health records

Machine learning techniques have many benefits over CAMD, but at the same time, they also have some limitations that must be addressed and solved. The decision error obtained by these techniques is smaller than the one of a doctor. The goal of this research is to reduce this error as much as possible. This means that this goal can be divided into two directions. The first refers to the increase of the effectiveness of the current algorithms and the second direction regards developing new algorithms, more performant than the existing ones.

By reducing the error of ML techniques we can also drop the cost for some hospitals. ML can be used as useful support for doctors in confirming the assumption regarding a possible disease. This can happen when a doctor is not 100% convinced about a certain diagnosis. Instead of carrying out additional investigations, a doctor can use the confirmation provided by the algorithm which also optimizes the flow of patients. Thus, some expensive medical tests can be replaced or are no longer needed. At the same time, for critical situations, ML techniques can improve the time needed to make a decision as it automatically ranks the medical data. Complex automatic decision systems must include ML algorithms to provide diagnosis instantly. This research proposes competitive decision systems whose purpose is integration with medical equipment.

It is worth mentioning that the novelty and effectiveness of the proposed results have been confirmed by publishing them in peer-reviewed international journals/proceedings. The models proposed will be presented in ascending order of importance to give a clear view of the main contribution of this research.

Algorithms hybridization. This refers to improving the performance result by merging two methods. The evolutionary paradigm is the one influencing the hybridization process. We hybridized a CNN with a genetic algorithm and this approach proved to have better performance than other similar techniques. A way of encoding a DL architecture in a chromosome was proposed and then different evolutionary operators were applied. The architecture of the build model was the best-selected chromosome. Our approach outperformed the classical DL algorithms which means that DL architectures can be successfully optimized by neuroevolution.

The evaluation process of the algorithms used to produce new ones. There are two approaches. First is the *statistical benchmarking process* which refers to individually evaluating an algorithm in the first phase and then comparing its performance with that of competitors. To address this problem, a mix of statistical comparison tests can be used. For comparing means and proportions we can use the two-sided *t* test and two-sided *z* test respectively. The *t*-test can be used for independent samples while the Mann-Whitney *U* test can be used as its non-parametric alternative. To verify the corresponding assumptions for the *t*-test, Kolmogorov-Smirnov/Lilliefors and Shapiro-Wilks tests can be used as normality tests while Levene's test/T-test can be used for homogeneity of variances.

The second approach is *the statistical-based individual assessment* which consists of using a set of quality measures to highlight the characteristic of each competitor's algorithm.

The statistical parameters refer to [6]:

- *Classification accuracy* - the portion of correctly classified cases in the training and testing stages.

- *Area under ROC curve* - the probability for a classifier to rank a chosen positive instance higher than a chosen negative one.
- *Sensitivity* - the percentage of correctly identified sick people as having the disease.
- *Specificity* - the percentage of correctly identified healthy people as not having the disease.

Medical database adoption for ML algorithms. The most common medical database consists of numerical data and medical digital images and movies. An ML technique requires numerical data so the challenge here is to 'adapt' the medical database from its form to a numerical form, but without losing the embedded data. A methodology which uses exploratory data analysis was proposed to accomplish this task. In addition, the vectorized versions of the digitized medical photos and movies, which had previously been converted into numerical matrix form through the hue histogram, served as the algorithm's input.

In 1998, a film digitized mammography system was approved as the first CAMD device and till now the number of such products has increased. These products play a key role in the healthcare sector, offering support for the most important disease detection. However, such products sometimes represent a challenge for doctors, despite the high technology embedded. The first reason is represented by the images provided by these instruments which are not high-definition. The second one is represented by the fact that the scanned images refer to moving images of organs which make the interpretation of the image quite difficult. Under these circumstances, sometimes it can be very difficult to interpret the images and thus subjectivity can intervene. To face this possible problem, integrating ML techniques in CAMD is a good solution.

This research is a result of the challenge of using ML algorithms to enhance the processing of medical data so that the medical diagnosis process can be improved. Starting from the standard algorithms, we tried to improve them and develop new ones. The focus was on colon cancer, lung cancer, breast cancer and brain cancer. The results were all validated by publishing them in well-established journals in Informatics, thus certifying their originality and novelty.

1.2. Achievements and restraints

The idea behind this research has been presented so far. Technically speaking, new ML algorithms are based on a two-step process:

- *Designing the algorithm* based on prior knowledge
- *Algorithm validation.* In order to obtain good performance, the new models' parameters have been set, using several imaging datasets and medical numerical datasets. The newly developed algorithms can be used with other datasets, be they medical or not. A suite of benchmark problems is used to verify the potential of the new algorithm. To validate the techniques for real-world problems, they have been compared to alternative approaches.

Regarding the constraints of this research, taking into consideration that we covered a few classical ML models, there remain many open problems that can be addressed in future work. There are many ML techniques used nowadays which means that we have to expand this work and consider new algorithms together with new real-world tasks in the CAMD domain.

1.3. Book structure

This book is structured into six chapters with the last one presenting achievements and final remarks. The first chapter is an introductory one, highlighting the need for intelligent decision-making systems in the healthcare sector. It gives a clear view of this research contribution and discusses the importance of designing and building new ML techniques that should be included in those systems to assist medical staff in making fast and accurate decisions.

Chapter 2 presents the state of the art in artificial intelligence methods in CAMD. It provides an overview of machine learning and deep learning techniques. The novelty of this research can be better evaluated based on this information.

Chapter 3 presents the medical datasets used to validate the proposed methods along with descriptive statistics and statistical techniques for evaluating algorithms performance. The datasets are both imaging datasets and numerical datasets. There are 5 medical imaging datasets: one concerning lung cancer, one regarding colon cancer, one concerning fetal brain ultrasound images, and two maternal-fetal datasets. In addition to these imaging datasets, another 4 numerical medical datasets have been used regarding lung cancer, colon cancer, ovarian cancer and breast cancer. One of the maternal-fetal medical dataset was created within an exploratory research project, namely "*Pattern Recognition and Anomaly Detection in fetal morphology using Deep learning and Statistical learning*". Some of the datasets required preprocessing in order not to impact the results.

Chapter 4 presents a good practice study of how to build a trustworthy intelligent system to improve patients' health. This study regards multi-doctor plus AI decision support and highlights two key features needed when working with doctors: collaborative systems and statistics used to validate both doctor's decisions and AI model's decisions. Thus, the doctor's performance, as well as the AI model's performance, are raised.

Chapter 5 presents the new AI techniques and models developed during this research. Inspired by evolutionary computation, a new approach to optimize a deep learning neural network architecture is proposed. Its effectiveness is demonstrated by comparing it with similar techniques. To address the random initialization of the hidden node parameters of an extreme learning machine, a new solution, which uses the Bayesian paradigm as an alternative, is presented. The experimental results showed the proposed variant is superior to that of the comparison models. Another novel approach is the development of 3 single hidden layer feedforward neural networks which initialize their input hidden layer weight matrix based on non-parametric correlation coefficients computed between the input and the output. Regarding robustness and performance, all the 3 models are comparable with state-of-the-art NNs.

Chapter 6 draws the conclusions and briefly summarizes the achievements of this research.

2. SOTA AI algorithms applied on medical datasets

2.1. Introduction

Nowadays, technology has a huge impact in many areas, especially in the medical field. A key role is represented by AI which can incorporate intelligence into different machines and makes them suitable for building smart systems for the healthcare sector. To build such systems there is a high need for new effective AI models. However, sometimes the task of developing an AI model can become challenging due to data variation diversity in real-world problems. This chapter provides a view on the state of the art in AI techniques and medical diagnosis applications.

2.2. Artificial Intelligence

There is a wide variety of processes behind the ability to learn. Regarding this action, there are many definitions we hear every day, such as "skill acquired by study" or "modification of a behavior by experience". When we hear these definitions, we think of human actions, but recently the challenge is to apply "learning" to machines.

The simulation of human intelligence by machines is called *Artificial Intelligence* (AI). There are different approaches that historically define the field of AI which go far back to Greek mythology. Its first definition in 1956 was "*the science and engineering of making intelligent machines*" according to John McCarthy [7]. It was the year the first AI laboratory was established. Now, AI dominates areas such as Expert Systems, Computer Vision or Deep Learning.